

ОТСЛЕЖИВАНИЕ ПОИСКОВЫХ РОБОТОВ

В.В. СТРЕЛЬЦОВ^{1*}, Е.П. ЧЕРНЫХ²

¹ студент кафедры ВТП, НТУ «ХПИ», Харьков, УКРАИНА

² доцент кафедры ВТП, НТУ «ХПИ», Харьков, УКРАИНА

* email: strelcov-v92@yandex.ru

Объем интернет аудитории растет с каждым днем. В настоящее время в глобальной сети можно найти множество способов и предложений для продвижения сайта. К наиболее значимым методам относятся поисковые системы.

«Веб-пауки» или поисковые роботы – программы, являющиеся составной частью поисковой системы. Поисковый робот является программой браузерного типа. Их принцип действия аналогичен обычным браузерам и предназначены они для перебора страниц. Заходя на тот или иной ресурс, они обходят все страницы сайта, либо часть страниц, после чего отправляют полученную информацию в поисковый индекс. В дальнейшем полученная и обработанная информация появляется в выдаче поисковых систем. Поисковые роботы могут обходить не все страницы, а лишь часть их, из-за чего могут возникнуть проблемы с индексацией больших сайтов. Похожие проблемы возникают по вине плохого хостинга, в результате чего часть страниц оказывается недоступной для анализа. Немаловажную роль играет наличие грамотной карты сайта и правильная настройка файла robots.txt. Из всех поисковиков самой большой коллекцией ботов располагает Яндекс. За индексацию отвечают следующие боты: главный индексирующий робот, собирающий данные со страниц web-сайтов; бот, способный распознавать зеркала; поисковый робот Яндекс, осуществляющий индексацию картинок; робот, сканирующий иконки favicon; несколько пауков, определяющих доступность страниц сайта. Главный поисковый робот Google занимается сбором текстовой информации. В основном он просматривает html-файлы, с определенной периодичностью анализирует JS и CSS. Способен воспринимать любые типы контента, разрешенные к индексации.

Для решения описанной выше проблемы был предложен подход – разработка программного модуля (скрипта), с помощью которого можно отследить визиты поисковых роботов. Скрипт будет содержать необходимую информацию: дату посещения, имя бота, IP-адреса бота и страницы, которые он посетил.

Программная реализация может быть выполнена с помощью языков программирования PHP или J.

Предложенный подход позволит повысить эффективность индексации сайта.